Promoter and intron 1 polymorphisms of COL1A1 interact to regulate transcription and susceptibility to osteoporosis

Huilin Jin, Rob J. van’t Hof, Omar M.E. Albagha and Stuart H. Ralston*

Rheumatic Diseases Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK

Received February 14, 2009; Revised April 9, 2009; Accepted April 28, 2009

Three polymorphisms (−1997G/T; −1663IndelT and +1245G/T) have been identified in the 5′ flank of COL1A1 gene that are associated with osteoporosis but the underlying mechanism is unclear. Here we investigated the functional effects of these variants on COL1A1 transcription. Transcription was 2-fold higher with the osteoporosis-associated G-del-T haplotype compared with the common G-Ins-G haplotype. Gel shift assays showed that the region surrounding the −1663IndelT polymorphism recognized a complex of proteins essential for osteoblast differentiation and function including Nmp4 and Osterix, and the osteoporosis-associated −1663delT allele had increased binding affinity for this complex. Chromatin immunoprecipitation assays confirmed that the region flanking −1663IndelT bound a complex of proteins including Osterix and Nmp4 and also showed evidence of recruitment of Nmp4 to the Sp1 binding site in intron 1. Further studies showed that haplotype G-del-T had higher binding affinity for RNA polymerase II, consistent with increased transcription of the G-del-T allele and there was a significant inverse association between carriage of G-del-T and bone mineral density (BMD) in a cohort of 3270 Caucasian women. We conclude that common polymorphic variants in the 5′ flank of COL1A1 regulate transcription by affecting DNA–protein interactions and that increased levels of transcription correlate with reduced BMD values in vivo. This is consistent with a model whereby increased COL1A1 transcription predisposes to osteoporosis, probably by increasing production of the alpha 1 chain and disrupting the normal ratio of collagen type 1 alpha 1 and alpha 2 chains.

INTRODUCTION

Osteoporosis is a common disease characterized by low bone mass, micro-architectural deterioration of bone tissue and enhanced bone fragility leading to an increased incidence of fracture. Genetic factors are recognized to play an important role in regulating bone mineral density (BMD) and fracture risk (1). Several candidate genes have been implicated including the COL1A1 gene that encodes the alpha 1 chain of type 1 collagen. Most studies of COL1A1 have focused on a polymorphism that affects a Sp1 binding site in the first intron at position +1245 (rs1800012) (2–7). Previous research has shown that the Sp1 polymorphism is a functional variant that affects DNA binding affinity, collagen transcription and collagen protein production, leading to disruption of the normal 2:1 ratio of collagen type 1 alpha 1 and alpha 2 chains (4). More recently, two polymorphisms have been identified in the proximal promoter of COL1A1 at positions −1997 (rs1107946) and −1663 (rs2412298) that are in linkage disequilibrium with the Sp1 polymorphism (8). These polymorphisms have been associated with BMD in several studies (9–12) and in some populations have been reported to interact with the Sp1 polymorphism to regulate BMD (8,12,13). Previous studies have shown that the region surrounding the −1663IndelT polymorphism binds the transcription factor Nmp4 and reporter assays have shown that promoter haplotypes affect transcription (14). It is currently unclear whether the promoter polymorphisms and the intron 1 polymorphism form an extended haplotype that regulates COL1A1 transcription however. In this study,

*To whom correspondence should be addressed. Tel: +44 1316511037; Fax: +44 1316511085; Email: stuart.ralston@ed.ac.uk

© The Author 2009. Published by Oxford University Press. All rights reserved.
For Permissions, please email: journals.permissions@oxfordjournals.org
we conducted functional analysis of all three polymorphisms and haplotypes in the 5' flank of COL1A1 in relation to their effects on gene transcription, transcription factor binding and bone density.

RESULTS
Regulation of reporter gene expression by 5' COL1A1 haplotypes
The strategy for generating the reporter vectors used to study the effects of COL1A1 haplotypes on transcription is shown in Figure 1 and the effects of the individual haplotypes on reporter gene expression, corrected for transfection efficiency are shown in Figure 2A. All of the reporter constructs had significantly higher expression of luciferase than the pGL3-basic vector confirming that the fusion construct produced a functionally active luciferase protein. The -1997G/-1663delT/+1245 T allele (haplotype 2; G-del-T) drove the highest levels of reporter gene expression, and this was significantly different from all the other haplotypes (P < 0.01). The -1997T/-1663insT/+1245 G allele (haplotype 3; T-ins-G) had the second highest luciferase activity, and this was significantly different from the -1997T/-1663delT/+1245 G allele (haplotype 7; T-del-G; P = 0.020), and the -1997T/-1663insT/+1245 T allele (haplotype 8; T-ins-T; P = 0.046). There was no significant difference between levels of reporter gene expression for the other haplotypes studied.

In order to determine if levels of gene expression were influenced by haplotype or by the genotype at individual polymorphic sites, we performed a further analysis in which gene expression was analysed on the basis of alleles at the individual polymorphic sites (Fig. 2B). For this analysis, we measured levels of expression for each construct and calculated average values depending on the allele present at each polymorphic site. This showed that there was no significant effect of individual polymorphisms on transcription indicating that the three polymorphisms interact with each other to form a haplotype that regulates transcription. The frequencies of the different haplotypes in the Aberdeen Prospective Osteoporosis Screening study (APOS) population (13) of white Caucasian women are shown in Figure 1C.

Association between transcriptional activity in vitro and bone density in vivo
We next investigated the relationship between the luciferase expression and BMD at the lumbar spine and femoral neck in 3270 post-menopausal women from the APOS study who had already been genotyped for these polymorphisms as previously reported (13). For this analysis, we categorized participants on the basis of whether or not they were carriers of the -1997G/-1663delT/+1245 T (G-del-T) allele, which was associated with the highest levels of luciferase expression. The results are shown in Figure 3 which shows that BMD values adjusted for age, age at menopause and body mass
The region surrounding \(-1663\text{delT}\) binds the transcription factors Nmp4 and Osterix

Previous studies have shown that the polyT tract surrounding the \(-1663\text{delT}\) polymorphism binds the transcription factor Nmp4. Examination of the sequence surrounding this site also revealed the presence of a binding site for the transcription factor Osterix (tgctctccccctctgctc; Professor Benoit De Crombrugghe, unpublished data). We therefore wanted to investigate whether Osterix binds to the DNA surrounding \(-1663\) and to confirm previous reports of Nmp4 binding. There was a shifted band using nuclear extracts from TE85 osteoblast-like cells with the \(-1663\) probe and supershift assays indicated that the protein complex contained Nmp4, Osterix and \(\beta\)-actin (Fig. 4A). We went on to perform competition assays in the electrophoretic mobility shift assay (EMSA) using excess amounts of unlabelled competitor probes and this showed that the \(-1663\text{delT}\) allele competitor probe had a significantly \((P < 0.001)\) greater affinity for protein binding than the \(-1663\text{insT}\) allele. (Fig. 4B). Quantitation of binding from three independent experiments showed a significant difference in binding affinity for the \(-1663\text{delT}\) and the \(-1663\text{insT}\) alleles (Fig. 4C).

Similar results were observed when the \(-1663\text{insT}\) allele was used as the labelled probe and the delT allele as the competitor (data not shown). The gel shift assays described above were repeated with nuclear extracts from MG63 cells with similar results (data not shown).

The region surrounding \(-1997\text{G/T}\) binds unidentified nuclear proteins

Examination of the sequence surrounding the \(-1997\text{G/T}\) polymorphism revealed consensus binding sites for Sp1 and Usf and in view of this, EMSAs were performed to determine whether the DNA surrounding the \(-1997\text{G/T}\) site bound these proteins. Addition of recombinant human Sp1 protein to the \(-1997\text{T}\) probe resulted in a single shifted band (Fig. 5A, Lane 2), but there was no shift using recombinant Usf protein (Fig. 5A, Lane 3). This indicates that the DNA surrounding \(-1997\) binds Sp1 but not USF \(\text{in vitro}\). Addition of TE-85 nuclear extracts to the probe resulted in a different banding pattern from that of purified Sp1 (Lane 4) and neither

Figure 2. Regulation of reporter gene transcription by COL1A1 haplotypes. (A) Luciferase expression of COL1A1 haplotypes H1-H8, corrected for transfection efficiency. Values are expressed as the percent change relative to the pGL-3 basic vector which is set at 100%. (B) Effect of individual polymorphisms on luciferase expression, analysed by categorizing vectors into series according to the allelic differences at each polymorphism. No significant difference was observed between alleles at the individual sites. The data shown in A and B are derived from four independent experiments. The columns are means and bars are SEM. **\(P < 0.01\); *\(P < 0.05\) from other vectors. (C) A schematic representation of the haplotypes studied and their frequency in APOSS population. The position of the polymorphisms is shown in relation to exons 1–6 of COL1A1 which are indicated by black bars.

Figure 3. Relationship between BMD and carriage of haplotype 2. (A) Bone mineral density values at the lumbar spine in the APOSS study population in relation to carriage of the G-del-T haplotype which had the highest luciferase expression \(\text{in vitro}\). (B) Bone mineral density values at the femoral neck in the APOSS study population in relation to carriage of the G-del-T haplotype. ++/++ homozygotes for G-del-T; +/- heterozygotes for G-del-T; -/- non-carriers of G-del-T. The columns are means and bars are SEM. The \(P\)-values are derived from ANOVA analysis across the three haplotype groups.
anti-Sp1 nor anti-Usf antibodies caused a supershift (Lane 5 and 6). Similar results were obtained when using a −1997G probe (data not shown). These results suggest that although purified Sp1 can bind the DNA surrounding the −1997G/T site, Sp1 does not appear to be one of the proteins responsible for DNA binding in nuclear extracts from osteoblast-like cells. Competition assay showed no consistent differences in binding affinity for nuclear proteins as reflected by attenuation of the DNA–protein complex more rapidly with increasing concentrations of the unlabelled delT competitor (Lanes 5–7) compared with the unlabelled insT competitor (Lanes 2–4). (C) Quantitative analysis of binding curves for the −1663insT and −1663delT competitor probes. Band intensities at three concentrations derived from three experiments were plotted as a percentage of band intensity in the absence of competitor for each competitor concentration. Symbols are means and bars are SEM. There was significant difference (P < 0.001) between the alleles.

Analysis of transcription factor binding using chromatin IP assays

Since EMSAs do not take account of the effects of chromatin on transcription factor binding, we went on to perform a chromatin immunoprecipitation (ChIP) assay to study transcriptional activation of the COL1A1 promoter. These studies were performed in MG63 cells, since it was found by DNA sequencing to be heterozygous for haplotype 2 (−1997G/−1663delT/+1245T) and 3 (−1997T/−1663insT/+1245G). Therefore, by using this cell line, it is possible not only to investigate DNA–nuclear protein binding, but also to quantify haplotype-specific binding by real-time PCR. These assays confirmed binding of Nmp4, Osterix and β-actin to the region surrounding the −1663insT polymorphism.
As expected, we confirmed binding of Sp1 to the region surrounding the +1245 polymorphism, but there was also evidence of Nmp4 binding to this region as reflected by a faint band on the IP assay with Nmp4 antibody (Fig. 6B). Since the region surrounding the +1245 polymorphism does not contain a consensus Nmp4 binding motif, this raises the possibility that, in the chromosomal context, Nmp4 protein bound to the −1663 site in the promoter may interact with proteins bound to the +1245 site in intron 1 to regulate transcription.

We went into perform quantitative PCR to determine whether the different haplotypes had differing binding affinity and whether there was a difference in transcriptional activity between haplotypes. For these experiments, we used an antibody to RNA polymerase II (RNAPII) as a surrogate for transcription and analysed expression of the different haplotypes using probes specific for haplotype 2 (+1245T) and haplotype 3 (+1245G) in the quantitative PCR (Fig. 6C). This assay showed a ~2.25-fold increase in binding of RNAPII to haplotype 2 compared with haplotype 3 ($P = 0.042$) which is in agreement with the findings in the reporter assays. The assay also showed a ~1.75-fold increase in Sp1–DNA binding affinity for the +1245 T allele of the Sp1 polymorphism, consistent with what was found in previous studies using EMSA (3).

**DISCUSSION**

Osteoporosis is a common disease with a strong genetic component characterized by reduced BMD and increased fracture risk, but the genes that regulate susceptibility are incompletely defined. Many candidate gene studies have been conducted in an attempt to identify alleles that predispose to osteoporosis but with a few exceptions (5,15,16), the studies have been underpowered and many reported associations have not been replicated. More recently, genome-wide association studies have been successful in identifying loci that regulate BMD and susceptibility to fracture (17–19). The genes implicated include *TNFRSF11A, TNFSF11* and *TNFRSF11B*, which encoding RANK, RANKL and osteoprotegerin, respectively; *ESRI*, which encodes estrogen receptor alpha; *SOST*, which encodes Sclerostin; *LRP5*, which encodes lipoprotein receptor related-protein 5; *SP7*, which encodes Osterix and the *MARK3* gene, which encodes MAP/microtubule affinity regulated kinase 3. However, the individual and cumulative effects of the variants identified so far are small and it is clear that the vast majority of genetic variants that regulate susceptibility to osteoporosis remain to be discovered. Although the *COL1A1* locus has not emerged as a significant determinant of BMD or fracture in recent genome-wide association studies, the Sp1 polymorphism and promoter polymorphisms which have been investigated here are not efficiently tagged by the SNP that have been used in the genome-wide association studies performed so far ($r^2 < 10\%$). Therefore, *COL1A1* remains a potentially important candidate for genetic regulation of susceptibility to osteoporosis and fragility fractures (5,13,20). The most intensively investigated polymorphism of *COL1A1* affects a Sp1 binding site within intron 1 at position +1245 relative to the transcription start site (2). The Sp1 polymorphism has been associated with BMD and vertebral fracture in many studies (4) and has been shown to be a functional variant with effects on DNA binding, gene transcription, protein production and mineralization of bone (4,7). More recently two polymorphisms have
been identified in the COL1A1 promoter at positions −1997 and −1663 which are in linkage disequilibrium with the intron 1 polymorphism (8). Previous association studies have suggested that the −1997G/T, −1663indelT and +1245G/T polymorphisms interact to regulate BMD (8,13). All three of these polymorphisms are in an evolutionary conserved region of COL1A1 but the mechanisms by which they regulate transcription and predispose to osteoporosis are incompletely understood.

In the present study, we conducted a functional analysis of the COL1A1 promoter using a combination of gel shift assays, promoter reporter assays and ChIP assays. The promoter-reporter assays showed that polymorphic variation at all three sites contributed to regulation of transcription and that the haplotype G-del-T (haplotype 2) had significantly higher transcriptional activity than all the other haplotypes. We also showed that carriage of haplotype G-del-T was inversely associated with BMD in a large scale population-based study (13), which is in agreement with findings reported by Bustamante et al. (22) in Spanish post-menopausal women. Although the effects of the extended COL1A1 haplotype on transcription have not been previously investigated, these observations are in broad agreement with previous studies, which showed that the promoter haplotype −1997G/−1663delT had high transcriptional activity (14). Taken together, these observations strongly support the hypothesis that the mechanism underlying the association of COL1A1 haplotypes with osteoporosis is increased transcription of the COL1A1 gene resulting in an abnormal ratio of α1(I) protein relative to α2(I) which adversely affects bone mineralization and bone strength by mechanisms partly independent of BMD (7,23,24). The importance of the haplotype rather than individual polymorphisms in driving these effects was emphasised by the fact that individual polymorphisms had no consistent effect on reporter gene activity. We previously reported that haplotype 6 (T-del-T) was over-represented in a small cohort of patients with hip fracture when compared with population-based controls (23), but in the present study, haplotype 6 was not associated with an increase in reporter gene expression. Possible explanations for this discrepancy are that the hip fracture association noted previously may have been a false-positive result or that haplotype 6 predisposes to hip fracture by a mechanism that does not depend on alterations in transcription. Further research will be required to clarify this issue, but it could be that haplotype 6 is in linkage disequilibrium with other functional polymorphisms in COL1A1 that affect the coding sequence or mRNA stability.

We also investigated the transcription factors that interacted with DNA surrounding the promoter polymorphisms using EMSA and ChIP assays. The EMSA experiments showed that the transcription factors Nmp4 and Osterix both bound to the DNA sequence surrounding the −1663indelT polymorphism. Moreover, antibodies to these proteins caused identical supershifts, suggesting that they form part of the same DNA/protein complex. It is very likely that additional proteins are present in the complex as we observed another super-shift by addition of anti-β-actin antibody. Although actin is classically considered to be a cytoskeletal protein, nuclear actin is known to bind to DNA and to participate in transcriptional regulation.
assays showed that in the context of Osterix/Nmp4/actin protein complex, whereas the reporter DNA, as depicted in Figure 7. The EMSAs reported here show promoter and intron 1 may directly interact through looping of sus motif for Nmp4. This suggests that the regulatory sites in the binding to the ChIP assays performed here showed evidence of Nmp4 cis bend or loop DNA, thereby facilitating communication between tion factors regulate gene expression by virtue of their ability to function. Two poly (dT) sequences in the rat

Figure 7. Suggested mechanism of osteoporosis associated with polymorphisms in the 5′ flank of COL1A1. Polymorphisms in the 5′ flank of COL1A1 regulate binding affinity of several critical nuclear binding proteins including Nmp4, Osterix and Sp1. This increases transcription of haplotype 2 which causes increased expression of COL1A1 mRNA, and over production of the alpha 1 chain relative to alpha 2, such that a proportion of collagen produced by osteoblasts is in the form of alpha 1 homotrimers. This adversely affects mineralization of bone, resulting in reduced BMD and an increased risk of fracture. (24,25). There is some evidence that actin participates in the regulation of COL1A1 transcription, since disrupting the actin cytoskeleton with cytochalasin B and microgravity has been shown to alter COL1A1 promoter activity (26). The fact that Osterix binds to the −1663 region is highly significant, since this transcription factor is known to be essential for osteoblast differentiation, bone formation and collagen production (27). The Nmp4 protein also plays an important role in osteoblast function. Two poly (dT) sequences in the rat Coll1a1 promoter region have already been identified as binding sites of Nmp4 (28) and mutation of either site increased promoter activity in rat osteoblast-like cells in reporter assays (29). The Nmp4 protein is one of a family of Cys2His2 zinc finger proteins that bind to the AT-rich minor groove of DNA. These transcription factors regulate gene expression by virtue of their ability to bind or loop DNA, thereby facilitating communication between cis-elements and trans-acting proteins (30). Interestingly, the ChIP assays performed here showed evidence of Nmp4 binding to the +1245 region, even thought this has no consensus motif for Nmp4. This suggests that the regulatory sites in the promoter and intron 1 may directly interact through looping of DNA, as depicted in Figure 7. The EMSAs reported here show that the −1663delT allele had increased affinity for binding the Osterix/Nmp4/actin protein complex, whereas the reporter assays showed that in the context of −1997G and +1245T, the −1663delT allele enhances DNA binding and transcription. It should be noted that Garcia-Giralt et al. (14) found that the −1663insT allele had increased affinity for Nmp4 binding, which contrasts with the results reported here. However, this apparent discrepancy is likely to be due to methodological differences since in the previous study, a single stranded probe was used which did not include the full Osterix binding site, whereas in this study we used a double-stranded probe which did include the Osterix binding site.

The role of the −1997G/T polymorphism in regulating COL1A1 transcription remains enigmatic. The −1997 polymorphism has been associated with BMD in several studies (8,11,13) and here we found that the DNA surrounding this polymorphic site contained a consensus binding site for Sp1. Although the −1997 region was found to bind purified Sp1 in vitro, there was no evidence of Sp1 binding in osteoblast nuclear protein extracts. This suggests that nuclear proteins other than Sp1 interact with the −1997 site in vivo. Although the polymorphism did not affect nuclear protein binding affinity, it could be that transcription factor complexes which bind to other regulatory sites in the gene interact with the −1997 complex to exert effects on its activity or stability.

Since EMSA and reporter assays do not take account of the effect of chromatin on transcription factor binding and gene expression, we performed further studies using ChIP assays to determine whether different COL1A1 haplotypes have differential effects on transcription. These experiments were performed in MG63 osteoblast-like cells since they were found to be heterozygote for haplotypes 2 and 3, and this allowed us to study allelic and haplotype-specific transcription factor binding and gene expression. To assess haplotype expression, we studied RNAPII binding as this has been shown to reflect the transcriptional activity of genes (31). This method can also be used for detecting the effects of SNP that do not contribute to RNA transcripts, as is the case for the promoter and intronic SNP studied here. These assays confirmed binding of Osterix and Nmp4 to the region surrounding the −1663delT polymorphism and showed a 2-fold increase in binding in haplotype 2 compared with haploype 3. This in agreement with our findings using the reporter assays. Moreover, haplotype 2 had higher binding affinity for Sp1 than haplotype 3 in MG63 cells consistent with the findings reported previously where the +1245 T allele was found to have higher DNA binding affinity than the G allele.

Taken together, the results presented here show that the three common polymorphisms in the 5′ flank of COL1A1 interact to regulate COL1A1 transcription by affecting binding affinity of several critical regulatory factors including Nmp4, Osterix and Sp1. The observations are consistent with the model depicted in Figure 7 whereby polymorphic variations in the 5′ flank inter-act to increase transcription of COL1A1. This, in turn, appears to reduce bone mass by altering the normal 2:1 ratio between collagen α1(1) and α1(2) chains, which subtly impairs mineralization of bone and increases fracture risk (4,7).

MATERIALS AND METHODS

Construction of COL1A1 reporter vectors

Reporter constructs comprising ~4 kb of the 5′ human COL1A1 sequence (−2310 to +1723), including the promo-
ter, exon 1, intron 1 and part of exon 2, were cloned upstream of the coding sequence of firefly luciferase in the pGL-3 vector (Promega). The 4 kb COL1A1 fragment was generated by PCR using Taq DNA polymerase (Qiagen) with the following primers: TCACAGCCAGGTACCCAAA (forward) and TCACGTACATCGCAACACCT (reverse). The PCR reaction consisted of an initial incubation for 5 min at 94°C, followed by 9 cycles of 15 s at 94°C, 1 min at 95°C and 4 min at 68°C, followed by 20 cycles of 94°C for 1 min, 56°C for 1 min and 68°C for 4 min with a 10 s increase per cycle. The PCR product was cloned into the Kpn I site of the pGL-3 basic luciferase vector, as shown in Figure 1. We generated all eight possible haplotypes defined by the three polymorphisms by site-directed mutagenesis using the Quickchange® II XL kit (Stratagene, UK). Identity of the constructs was verified by DNA sequencing. To confirm that the COL1A1–luciferase fusion construct mRNA was faithfully expressed, total cell RNA was extracted and reverse transcribed using the SuperScript III Reverse Transcriptase kit (Invitrogen). The sequence surrounding the COL1A1 and luciferase sequences was amplified by PCR using a forward primer in COL1A1 exon 1 (5’-ATGTTCAGCTTTGTGGACCT-3’) and the reverse primer in the luciferase sequence (5’-AGTTGCTCTCCACGGGTT-3’) using the following cycling protocol: initial denaturation 95°C for 10 min; followed by 35 cycles of denaturation 95°C for 15 s, annealing 60°C for 30 s, extension 72°C for 15 s. The PCR products were sequenced in both directions using the PCR primers as the sequencing primers.

### Cell culture, transfection and measurement of luciferase activity

The constructs were transfected into TE-85 human osteosarcoma cells cultured in alpha MEM medium supplemented with 10% fetal calf serum, 50 IU/ml penicillin and 2 mM glutamine, using the Magnet Assisted Transfection method (IBA, Germany) according to the manufacturers instructions. The pRL-TK vector (Promega) was co-transfected with each construct to provide a control for transfection efficiency. Twenty-four hours after transfection, the cells were lysed and luciferase activity was measured by using the Stop and Glow luciferase assay (Promega). The levels of firefly luciferase were corrected by the signal from the TK-renilla control vector. Expression of the mRNA from the COL1A1–luciferase fusion protein construct was confirmed by reverse transcription–PCR. For these experiments, total RNA was extracted from TE-85 cells which had been transfected with the constructs using RNA Bee (AMS Biotechnology) and reverse transcribed using the SuperScript III Reverse Transcriptase kit (Invitrogen).

### Electrophoretic mobility shift assays

Nuclear extracts from TE-85 and MG-63 cells were prepared by using a nuclear protein extraction kit (Active motif). HPLC purified 5′ biotin-labelled oligonucleotide probes were obtained from Invitrogen and double-stranded probes were made using the forward and reverse oligonucleotides as listed in Table 1. The EMSA binding reactions were performed at room temperature using the LightShift Chemiluminescent EMSA Kit (Pierce, USA) according to the manufacturer’s instructions. Competition EMAS were carried out to compare the binding affinities of the polymorphic variants for relevant nuclear proteins by adding unlabelled oligonucleotides of varying concentration. In the super-shift assays, antibodies were added to the binding reaction before adding the biotin-labelled probe. The antibodies used were as follows: Usf and Sp1 (Santa Cruz); β-actin (Sigma); Nmp4 (a gift from Professor Joseph P. Bidwell, Indiana University) and Osterix (a gift from Professor Benoit de Crombrugghe (Department of Molecular Genetics, Houston, TX, USA). Band intensity in the EMSA experiments was quantified by Genetools software (SynGene).

### Genotyping of osteoblast-like cell lines

The osteoblast-like cell lines TE85 and MG63 were grown to 90% confluence in alpha MEM medium supplemented with 10% fetal calf serum, 50 IU/ml penicillin and 2 mM glutamine. Genomic DNA was extracted using the DNAeasy Blood and Tissue kit (Qiagen) and a 4 kb fragment of COL1A1 comprising the promoter, exon 1, intron 1 and part of exon 2 was generated by PCR as described above. The PCR amplified fragments were cloned into pGL-3 basic vector and analysed by DNA sequencing to determine the haplotypes present in each cell line.

### Haplotype-specific ChIP

Haplotype-specific ChIP (HaploChip) assays were performed on MG63 cells. The ChIP assays were carried out as

---

**Table 1. Single-stranded oligonucleotides used for EMSA probe formation**

<table>
<thead>
<tr>
<th>Oligonucleotide</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1663insT (For)</td>
<td>Biotin-CGCGCAACGGACGTGCTCTCCCTCTTCTGTCTTCTTTTTTCCCTTTTGC</td>
</tr>
<tr>
<td>1663insT (Rev)</td>
<td>GCGGGGCGGGAAAGGGGAGGAAGGGGACGACGGTCGGCCG</td>
</tr>
<tr>
<td>1663delT (For)</td>
<td>GCGGGGGGAAAAGGGGAGGAAGGGGACGACGGTCGGCCG</td>
</tr>
<tr>
<td>1997G (For)</td>
<td>GGTCGTATTAGGAGGGGTTCTTGGAACTGACCC</td>
</tr>
<tr>
<td>1997G (Rev)</td>
<td>GGTCGTATTAGGAGGGGTTCTTGGAACTGACCC</td>
</tr>
<tr>
<td>1997T (For)</td>
<td>Biotin-CGCCTATTAGGAGGGGTTCTTGGAACTGACCC</td>
</tr>
<tr>
<td>1997T (Rev)</td>
<td>Biotin-CGCCTATTAGGAGGGGTTCTTGGAACTGACCC</td>
</tr>
<tr>
<td>Sp1 T (For)</td>
<td>Biotin-AGGAATGGCCGAGGATGAGCCC</td>
</tr>
<tr>
<td>Sp1 T (Rev)</td>
<td>Biotin-AGGAATGGCCGAGGATGAGCCC</td>
</tr>
</tbody>
</table>

The polymorphisms are underlined. For: forward oligo; Rev: reverse oligo.
essentially as described by Naughton (32). After cross-linking and sonication, chromatin was precleared with 1 μg anti-mouse IgG; 2 μg sheared salmon sperm DNA and protein-G-Agarose (50 μl of 50% slurry in dilution buffer) for 3 h at 4°C. Immunoprecipitation was performed overnight at 4°C with 2 μg sheared salmon sperm DNA, 50 μl protein-G-Agarose and specific antibodies. All antibodies were from the same source as mentioned in EMSA reactions except anti RNAPII (Covance). Precipitates were washed and reverse cross-linked followed by purification using QIAquick spin kit columns (Qiagen) and amplified using Taq DNA polymerase (Qiagen). PCR primer sequences for the −1997 polymorphism were: 5′ CCCGACCATGTTGCCAGCA 3′ (forward) and 5′ TCACCTACCTCATACCAAGC 3′ (reverse); and for the −1663 polymorphism were: 5′ TAGCCCTGCAGTCTCCCTC 3′ (forward) and 5′ AAGATTTCAATTGCCTCCCCC 3′ (reverse) and for the Sp1 polymorphism were: 5′ GATGTCTAGGTGCTGGAGGT- (forward); for the Sp1 (reverse) and 5′ GGCGAGGGAGGAGAGAAGG (reverse) and 5′ GGCGAGGGAGGAGAGAAGG 3′ (reverse). The PCR conditions for −1997, Sp1 and −1663 polymorphisms were previously described (13). The allelic/haplotype-specific binding was quantified by Taqman SNP genotyping assays and SensiMix DNA kit (Quantace) according to the manufacturer’s instructions.

**Statistical analysis**

Differences in luciferase expression across different haplotype types were assessed by analysis of variance (ANOVA) and the difference between individual haplotypes was assessed by Tukey’s test. Differences between BMD values in different haplotype groups were also assessed by ANOVA. Differences in DNA−protein binding affinity between alleles in the EMSA experiments were assessed by non-linear regression analysis. Differences in allele-specific transcription, as measured by quantitative PCR, in the ChIP assays was assessed by Student’s t-test. Statistical analysis of the EMSA data was carried out using GraphPad PRISM software assuming a one-tailed Student’s t-test. Statistical analysis of the EMSA data was carried out using GraphPad PRISM software assuming a one-tailed Student’s t-test.

**ACKNOWLEDGEMENTS**

We would like to thank Professor Benoit de Crombrugghe for helpful discussions and constructive comments on the manuscript and for kindly supplying the osterix antibody and the consensus sequence of the osterix binding site. We thank Dr Joseph P. Bidwell for kindly providing the Nmp4 antibody. We thank Dr Catherine Naughton, Dr Nick Gilbert and Dr Julian Knight for a number of useful discussions and suggestion in the HaploChIP assay.

**Conflict of Interest statement.** S.H.R. is an inventor of patents on the use of various genetic markers for the diagnosis of osteoporosis including the COLIA1 Sp1 polymorphism.

**FUNDING**

This study was supported in part by an academic support grant from the Arthritis Research Campaign (UK) and by a collaborative project grant from the European Commission (FP7-201865).

**REFERENCES**


